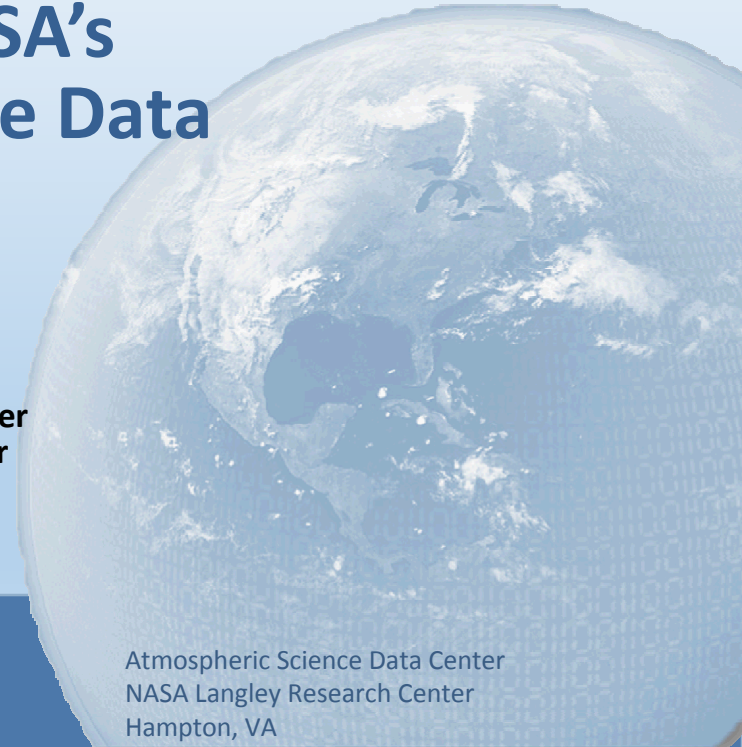


Openshift @ NASA's Atmospheric Science Data Center

Jeff Walter

**Atmospheric Science Data Center
NASA Langley Research Center
May 6, 2019**



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA



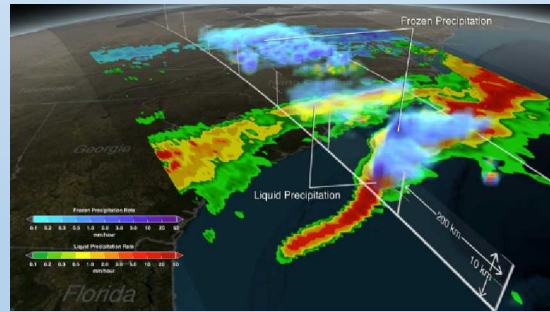
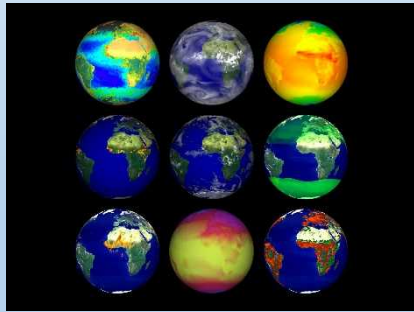
EXPLORE EARTH

YOUR HOME, OUR MISSION

NASA Earth Science Division

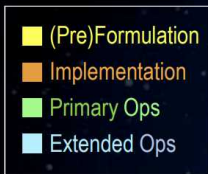
Key Science Questions

- How is the global Earth system changing?
- What causes these changes in the Earth system?
- How will the Earth system change in the future?
- How can Earth system science provide societal benefit?



Science Focus Areas

- Atmospheric Composition
- Weather
- Carbon Cycle and Ecosystems
- Water and Energy Cycle
- Climate Variability and Change
- Earth Surface and Interior
- Societal Benefit



NASA Earth Science

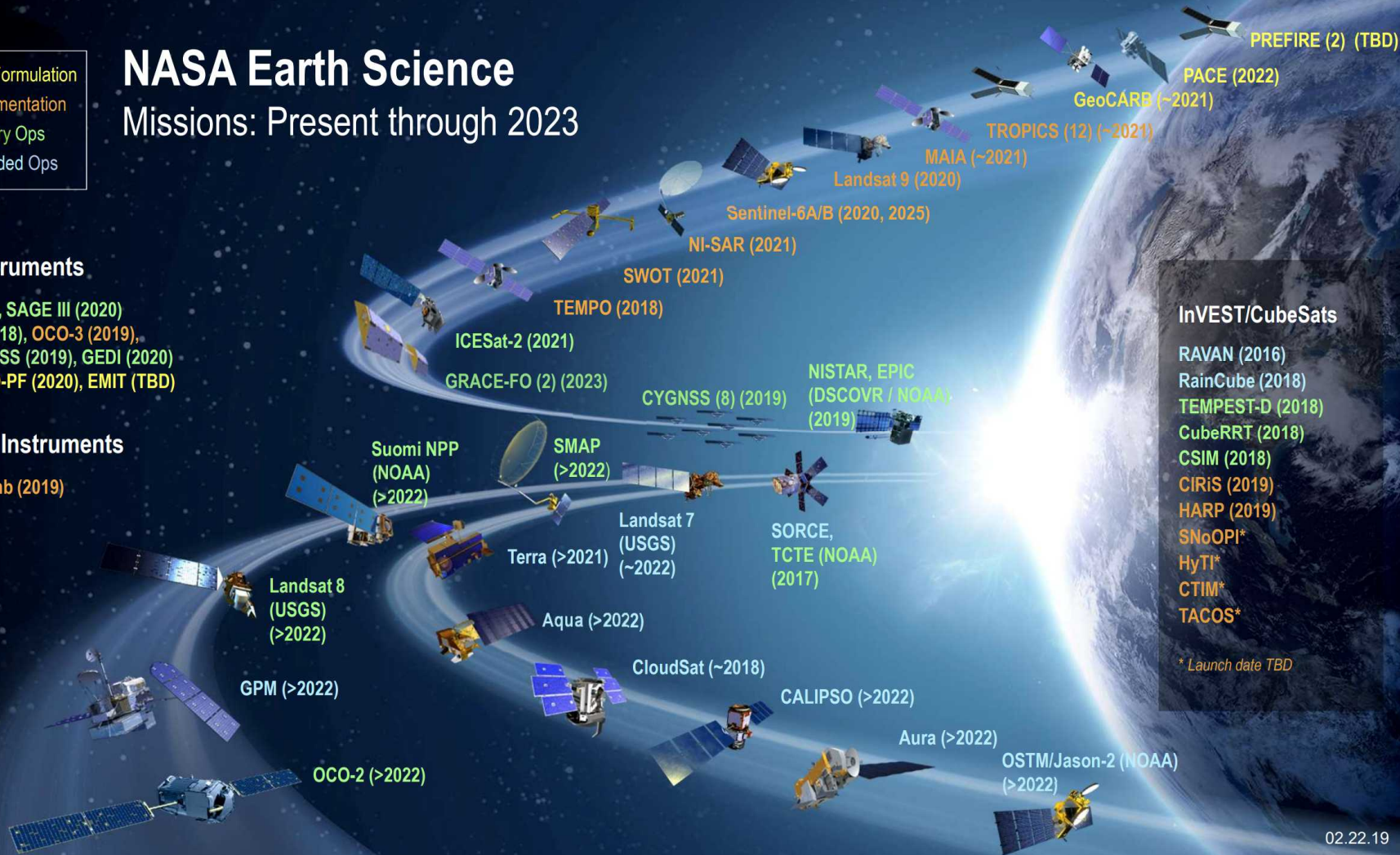
Missions: Present through 2023

ISS Instruments

LIS (2020), SAGE III (2020)
TSIS-1 (2018), OCO-3 (2019),
ECOSTRESS (2019), GEDI (2020)
CLARREO-PF (2020), EMIT (TBD)

JPSS-2 Instruments

OMPS-Limb (2019)

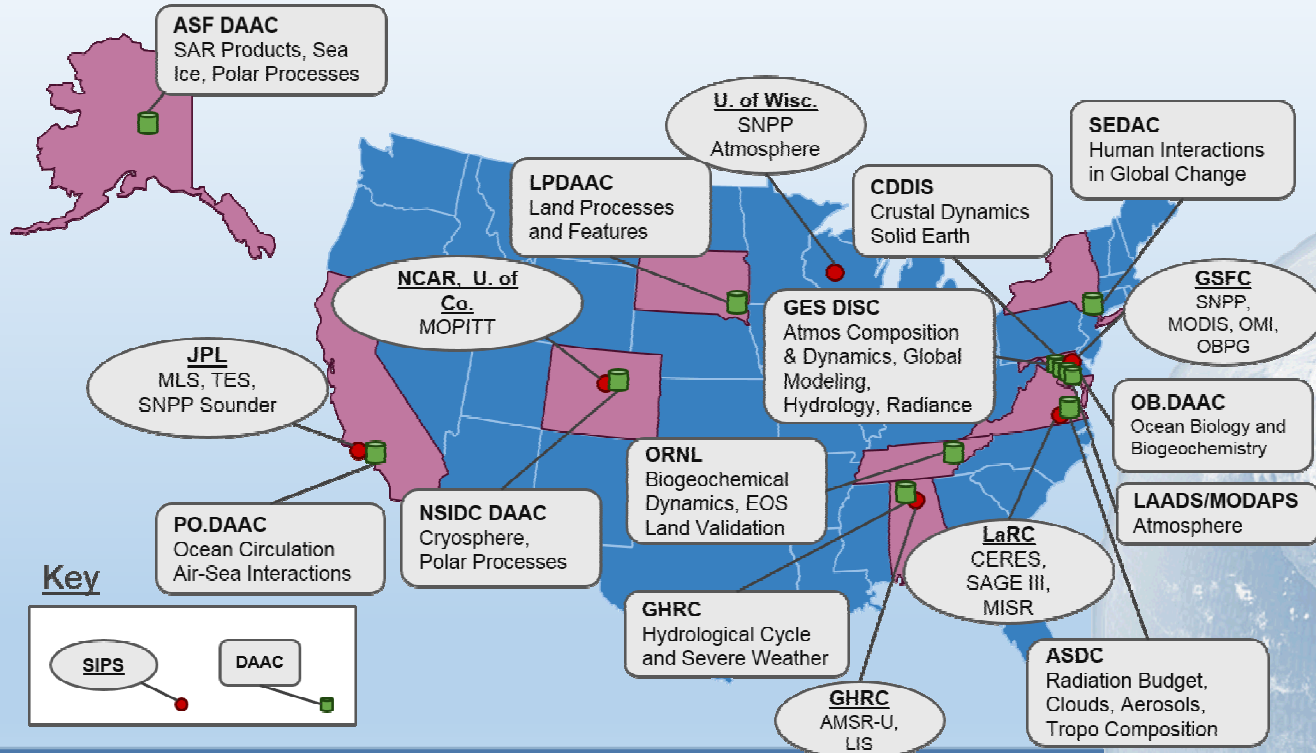


InVEST/CubeSats

RAVAN (2016)
RainCube (2018)
TEMPEST-D (2018)
CubeRRR (2018)
CSIM (2018)
CIRiS (2019)
HARP (2019)
SNoOPI*
HyTI*
CTIM*
TACOS*

* Launch date TBD

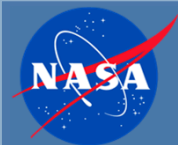
NASA Earth Science Data Systems Elements



Distributed Active Archive Centers (DAACs), collocated with centers of science discipline expertise, archive and distribute standard data products produced by **Science Investigator-led Processing Systems (SIPS)**

NASA Earth Science Data Policy

- NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available.
- There will be no period of exclusive access to NASA Earth science data. Following a post-launch checkout period, all data will be made available to the user community. Any variation in access will result solely from user capability, equipment, and connectivity.
- NASA will make available all NASA-generated standard products along with the source code for algorithm software, coefficients, and ancillary data used to generate these products.



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

What does ASDC do?

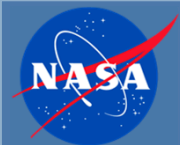
- Data ingest, archive and distribution services
- Operational science data production
- Science data production infrastructure hosting
- Scientific research computing infrastructure hosting
- Web hosting



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Challenges (aka Opportunities)

- Unsustainable physical complexity/stovepiping
- Challenging IT security and networking environment
- Multi-tenancy
- Aging storage
- Difficult to quickly innovate and add new functionality



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Strategic Objectives

- Reduce coupling between applications and the physical environment
- Reduce physical heterogeneity
- Eliminate stovepiping
- Improve reliability and robustness
- Improve our ability to rapidly respond to new requirements and deploy new functionality quickly
- Align with programmatic direction



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Openshift Projects

- Web Hosting Platform
- Ingest and Archive (Code Name: Parhelion)
- Science Data Production Management (Code Name: Hermes)



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

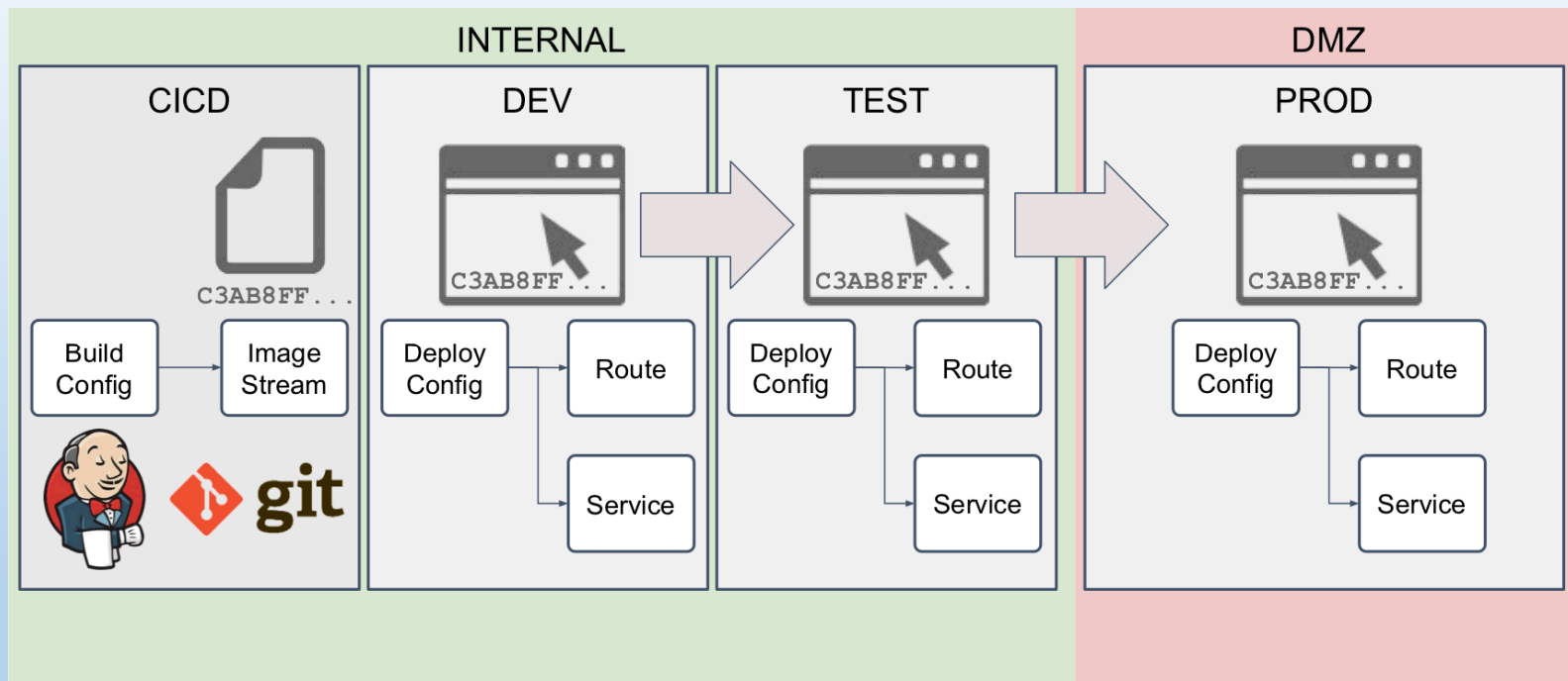
Web Hosting Platform (Current State)

- Hand-built FreeBSD jails
 - Manual updates/upgrades
- Hand-built VM servers
- Manual promotion between dev, test, production



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Web Hosting Platform (Future State)



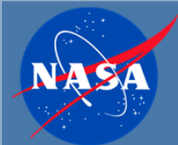
Data Ingest and Archive (Current State)

ANGe

- Created in house in circa 2007
- Monolithic
- Has served us well but...
 - Many one-off workflows
 - Many one-off submission interfaces
 - Technologically obsolete
- Challenging to maintain and operate
 - Heroic efforts
- Missing some key functionality
 - Prioritization

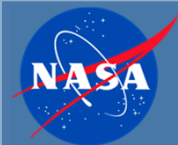
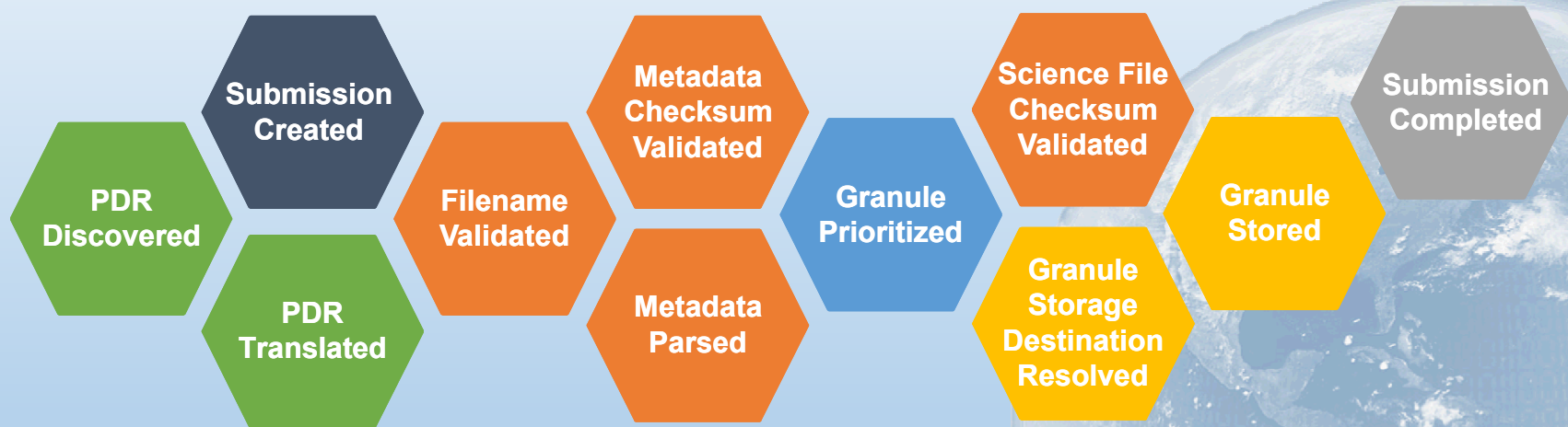
ECS

- Created in the late 1990's
- By millions of developers
- Billions of pages of documentation
- Costs program lots of \$\$\$ to maintain
 - They would love to retire it!
 - We can help!



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Data Ingest and Archive (Future State)



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Data Ingest and Archive (Future State)

Composable

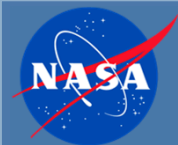
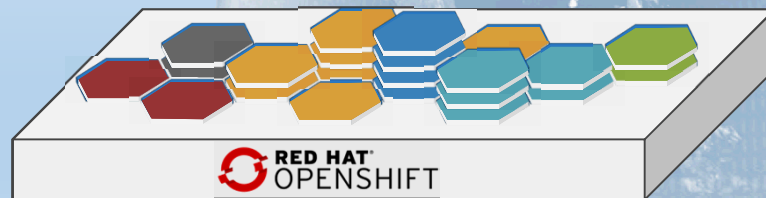
- Common Business Processes
- Loosely Coupled
- Linked together as needed
- Deployments of the “right” size
- Scale to meet demands



Services form a Workflow



Auto-Scale Services based on demand
when and where needed

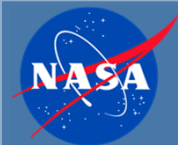
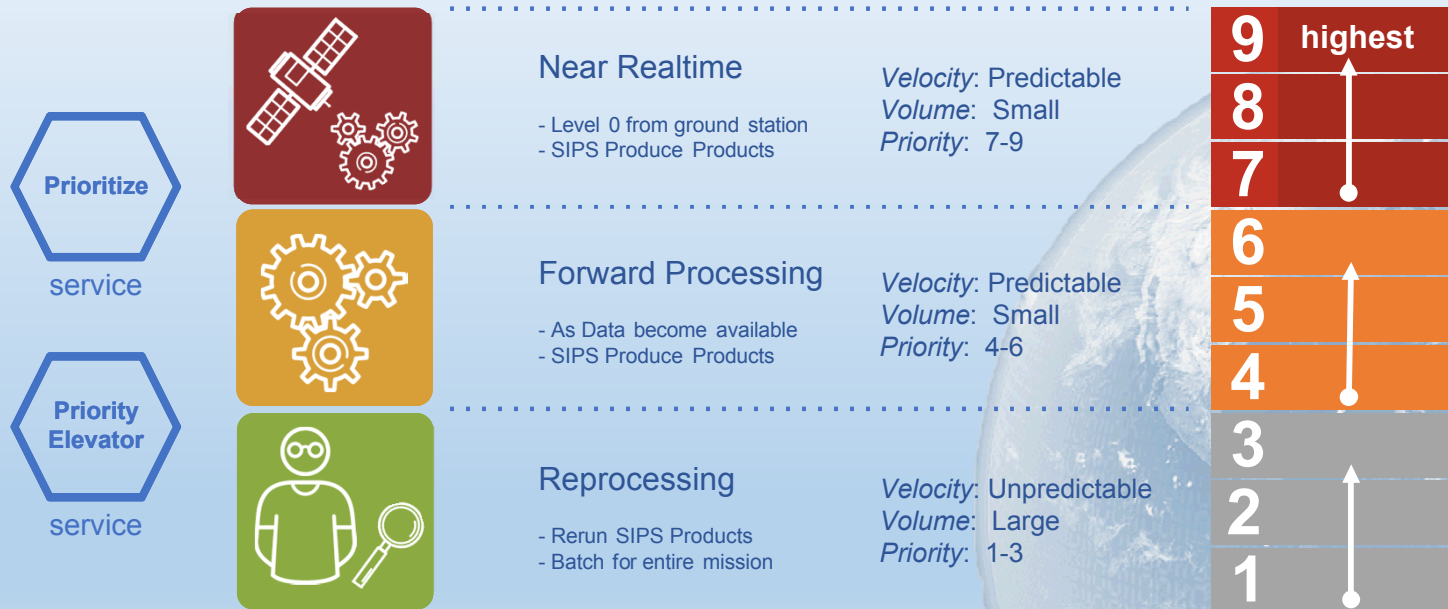


Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Data Ingest and Archive (Future State)

QUALITY OF SERVICE

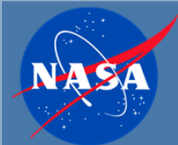
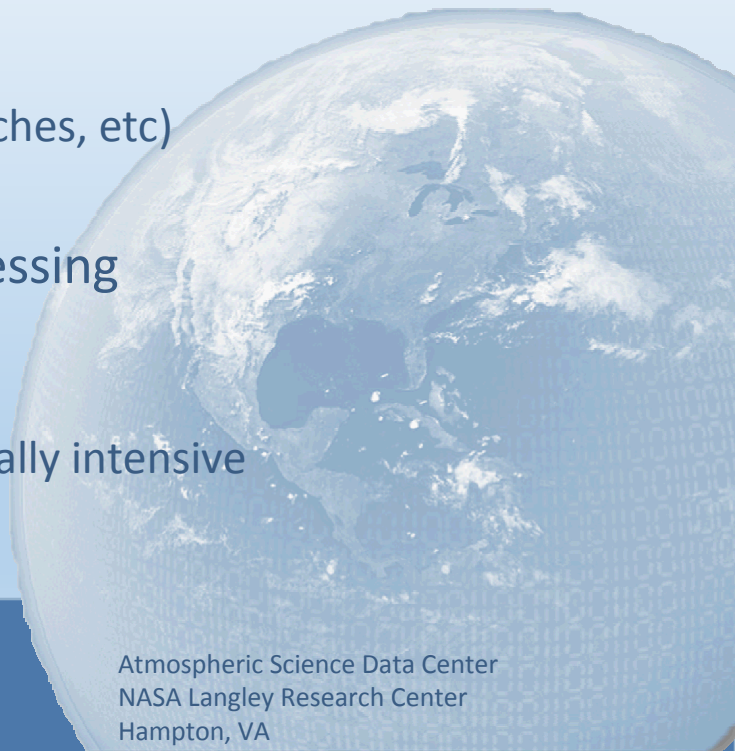
HIGHER PRIORITY SUBMISSIONS COMPLETE FIRST



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Science Data Production (Current State)

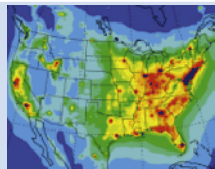
- Support multiple missions
 - Each has its own bare metal cluster
 - Single points of failure (head node, cluster switches, etc)
 - Unable to share idle hardware resources
- Different software systems to manage processing
 - Perl 5 based
 - Long in the tooth
 - Scenario configuration/reconfiguration is manually intensive
 - Difficult to track history and provenance



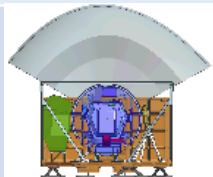
Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Multi-Angle Imager for Aerosols (MAIA)

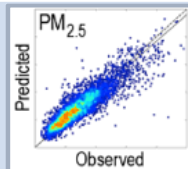
Within the System Scope



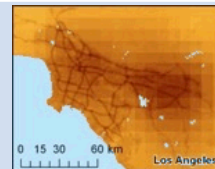
A chemical transport model (CTM) provides initial estimates of the abundances of different aerosol types.



The MAIA instrument uses multi-angle multi-spectral radiometry/polarimetry to retrieve aerosol optical depths (AOD_x) of different particle types over major population centers



Geostatistical regression models (GRMs) derived from collocated surface and MAIA measurements relate AOD_x to near-surface concentrations (PM_x).



Spatiotemporally gap-filled PM maps are generated by integrating PM surface monitor, MAIA instrument, and bias-corrected CTM data.



Geocoded birth, death, and hospital records and epidemiological methodologies are used to associate PM exposure with adverse health outcomes.

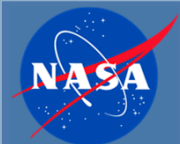
Credit: David Diner, JPL



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Science Data Production (Future State)

- Science algorithm executables containerized
- Batch job management with native Openshift capabilities
- Combine HPC and HTC approaches
- Graph database technology
- Resource sharing/accountability
- All science data production at ASDC managed by a single system instance



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

The Future

- Transition applications and systems to Openshift
- Collapse and virtualize infrastructure
- Explore transitioning archive to object storage
- Future-proof architecture
- Demonstrate the viability of a possible hybrid or multi-cloud future



Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA

Thank you!



Atmospheric
Science
Data Center

Atmospheric Science Data Center
NASA Langley Research Center
Hampton, VA